Supporting Information

Lenz *et al.* 10.1073/pnas.0804295105

SI Methods

SI Array CGH Statistical Methods

Array Segmentation and Copy Number Estimation. Each chromosome in each sample was divided into segments of similar log ratios according to the DNAcopy algorithm in bioconductor (www.bioconductor.org) (1), with alpha set to 0.1. For each array we plotted the average log ratio of each segment versus its length in probes [supporting information (SI) Fig. S1], identifying repeated long segments with similar average log ratios that likely represented different integer copy numbers. We modeled the segment averages according to the following formula (2) assuming the normal cells from the test and reference samples have copy number 2.

$$M_i = \log_2\left(\frac{2 + p(c_i - 2)}{2}\right) + d + \frac{\varepsilon_i}{\sqrt{n_i}}$$

where M_i, is the segment average, c_i is the true copy number for that segment, n_i is the number of probes in the segment, and ε_i is a mean 0 random error of the segment. The quantities p and d as well as the variance of ε_i are tuning parameters that were estimated separately for each array based on the plot of segment average versus segment probe number. Inverting this formula with the fitted tuning parameters, we arrive at the following estimate of copy number of each segment:

$$\hat{c}_i = 2 + \left(\frac{2^{M_i+1-d} - 2}{p}\right)$$

The normal reference DNA that was used was from a single male individual. To control for copy number variants in this reference, we performed hybridizations comparing this reference DNA with a pool of normal DNAs (Promega). This procedure identified short regions of copy number variation in the reference. Any segments in the experimental tumor samples that overlapped with 1 of these regions of copy number change in the reference were excluded from analysis.

Identification of Recurrent Abnormalities. We considered 4 classes of abnormal segments based on their estimated copy number

1. Single copy deletion ($\hat{c}_i < 1.5$)
2. Double copy deletion ($\hat{c}_i < 0.5$)
3. Gain of copy number ($\hat{c}_i > 2.5$)
4. Amplification ($\hat{c}_i > 3.5$)

Additionally a segment is called wild type, if ($1.5 < \hat{c}_i < 2.5$). For each of these 4 abnormality classes, we looked for recurrent abnormalities of each of the following 4 types:

1. Abnormal chromosome arm
2. Abnormal whole chromosome
3. Short recurrent abnormality (SRA)
4. Long recurrent abnormality

These types were derived as follows. A sample was defined as having an abnormal chromosome arm of a particular class (e.g., deletion or gain) if abnormal segments of that class covered more than 60% of the arm in that sample. An arm was defined as wild type if wild-type segments covered more than 95% of the arm.

A sample was defined as having an abnormal whole chromosome if it was found to be abnormal in both arms. A sample was defined as wild type for an entire chromosome if it was wild type for both chromosome arms.

SRAs were identified in a manner similar to the identification of minimal common regions (MCRs), as described previously (3). For illustration purposes, we describe how recurrent deleted regions were identified. The identical method was repeated for other classes of abnormality.

1. If 2 deleted segments on the same sample were separated by a gap shorter than 500 kb, and this gap was shorter than both the deleted segments on either side of the gap, then the gap was closed to generate a single longer deleted segment.
2. Deleted segments greater than 25 Mb were considered uninformative and were eliminated.
3. The chromosomal location that was covered by the largest number of deleted segments in different samples was identified. Those deleted segments were called the "overlapping group" of that location.
4. A deletion SRA then was defined by a core region that contained the chromosomal locations that were covered by at least two-thirds of the deleted segments in the overlapping group. An extended region was identified that consisted of the chromosomal locations that were covered by at least one-third of the deleted segments in the overlapping group.
5. A sample was declared to exhibit the deletion SRA if it was a member of the overlapping group. A sample was declared wild type for a SRA if wild-type segments covered more than 95% of the extended region.
6. Steps 1–5 were repeated, considering only locations that were not part of the extended region of any previously identified SRA and that included 2 or more samples in their overlapping group.

Long recurrent abnormalities were defined similarly, with the following exceptions. The algorithm for a long recurrent deleted region was

1. In step 1, all gaps < 10 Mb that were flanked by deleted segments at least 1.5 times their length were closed.
2. In step 2, segments < 15 Mb, as well as segments that were part of an abnormal chromosomal arm, were considered uninformative and were eliminated.

In our analysis, all 4 classes of abnormalities were combined under the term MCR, and the core region was used to define the extent of the MCR.

Refining Recurrent Abnormalities with Gene Expression. To define a list of recurrent abnormalities that influenced gene expression, we performed a permutation test as follows.

1. Consider all genes that are in the extended region of the recurrent abnormality or, in the case of arms and whole chromosomes, those that are located on the arm or on the chromosome.
2. For each gene calculate a 1-sided t test p-value for a difference in gene expression between the samples that exhibit the recurrent abnormality and those that are wild type for that abnormality, in the direction of increased gene expression being associated with increased copy number. If more than 1 probe set is available for a given gene, use the set that results in the lowest p-value.
3. Generate a statistic equal to the sum of the $-\log(p\text{-values})$ for the genes in the extended region.

4. Randomly permute the gene expression values and repeat steps 1–3 1000 times.
5. Consider only the MCRs for which the unpermuted statistic is $> 90\%$ of the same statistics calculated with the permuted data.

General Statistical Methods. The statistical significance of differences in MCR frequency between subtypes was calculated using a Fisher's exact test. Statistical significance for the relationship between survival and MCRs was calculated by comparing the survival of samples that exhibited an MCR with the survival of samples that were designated as wild type using a log-rank test.

SI Experimental Methods

Real-Time Quantitative PCR. Real-time quantitative PCR was used to evaluate copy number alterations detected by array CGH as described previously (4, 5). The genomic copy number of *SPIB* and *INK4a/ARF* was determined relative to the control genes *B2M*, and *PRKCQ* (Table S4).

Each assay was analyzed by the comparative cycle threshold method, using the arithmetic formula provided by the manufacturer. To determine the cut-off values for a genomic gain/amplification/single-copy deletion/homozygous deletion, 6 DNA samples from peripheral blood of healthy individuals were studied. The cut-off ratios for gain/amplification/single-copy deletion/homozygous deletion were determined as described previously (4, 5).

1. Venkatraman ES, Olshen AB (2007) A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* 23:657–663.
2. Fridlyand J, Snijders A, Pinkel D, Albertson D, Jain A (2004) Hidden Markov models approach to the analysis of array CGH data. *J Multivar Anal* 90:132–153.
3. Tonon G, *et al.* (2005) High-resolution genomic profiles of human lung cancer. *Proc Natl Acad Sci USA* 102:9625–9630.
4. Bea S, *et al.* (2005) Diffuse large B-cell lymphoma subgroups have distinct genetic profiles that influence tumor biology and improve gene-expression-based survival prediction. *Blood* 106:3183–3190.
5. Rosenwald A, *et al.* (2003) The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. *Cancer Cell* 3:185–197.

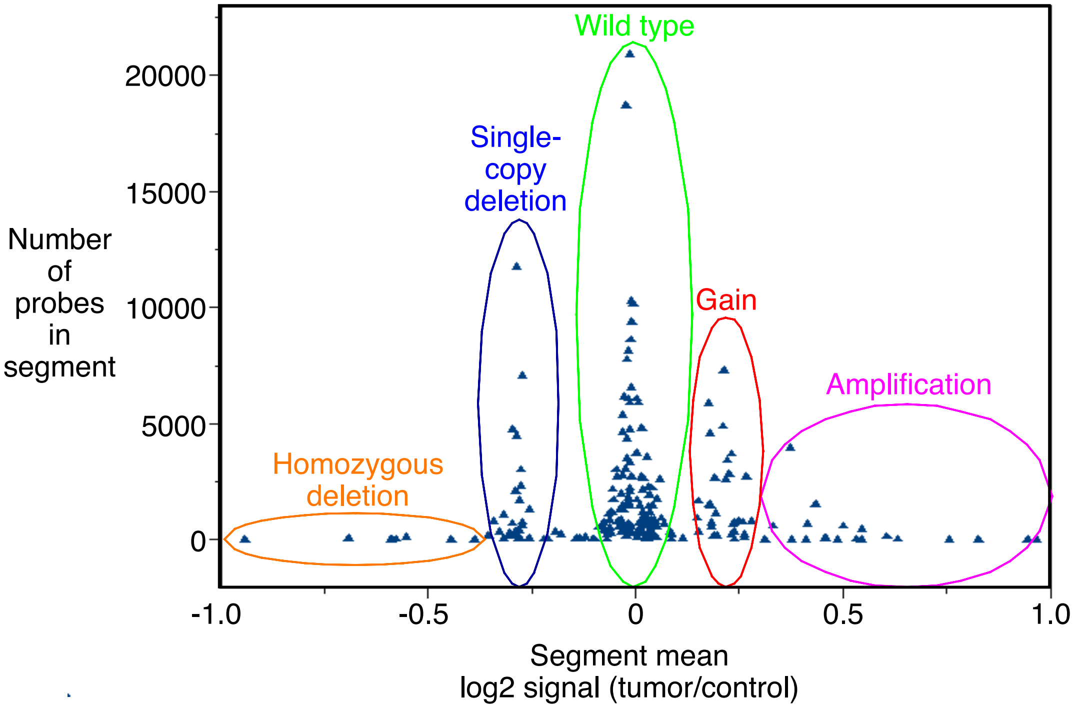

Fig. S1. Division of aCGH data into segments and classification of segments into single-copy deletion, homozygous deletion, single-copy gain, amplification, or wild type.

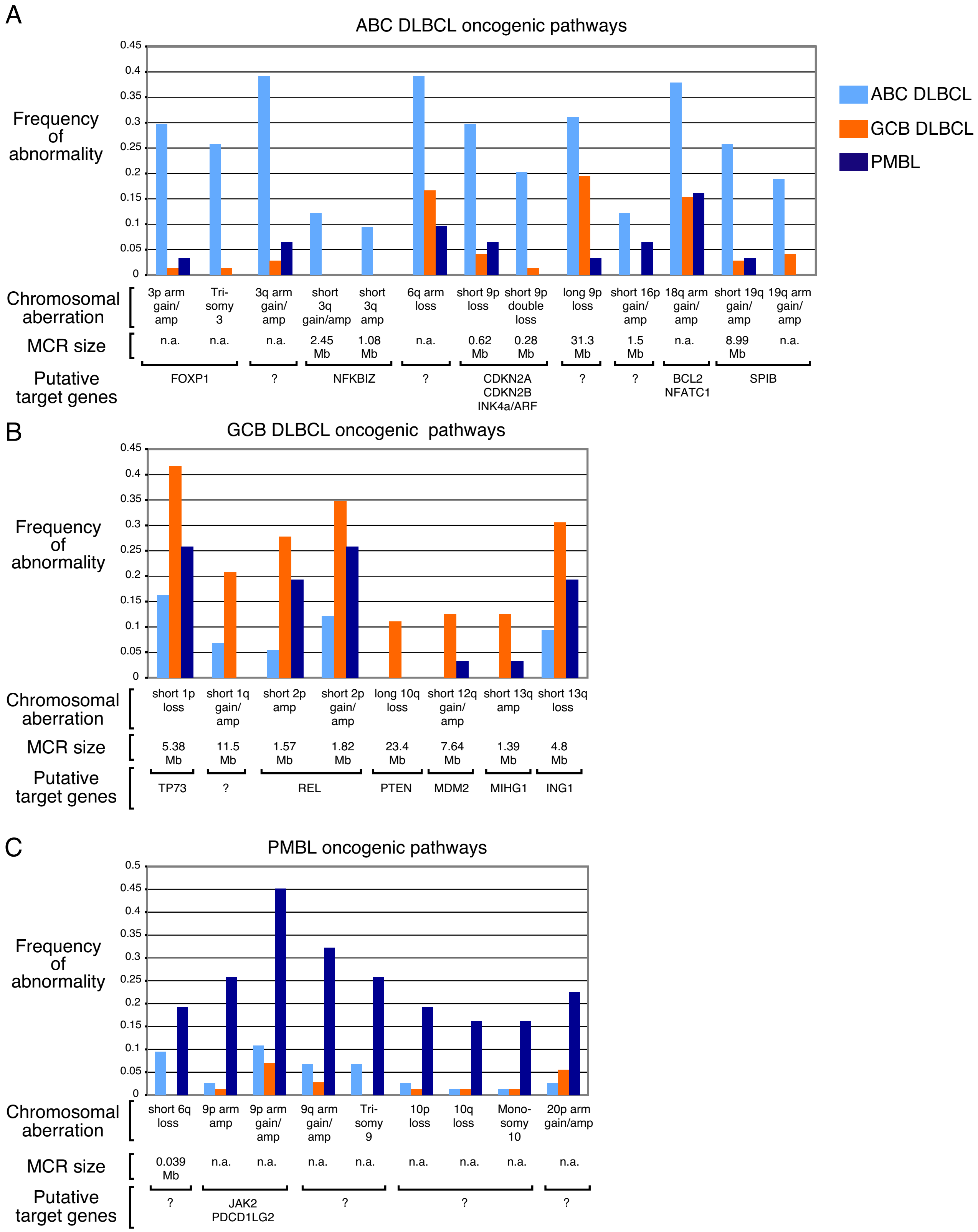

Fig. S2. (*A*) Complete panel of ABC DLBCL-specific genomic aberrations detected by aCGH. (*B*) Complete panel of GCB DLBCL-specific genomic aberrations detected by aCGH. (**C**) Complete panel of PMBL-specific genomic aberrations detected by aCGH.

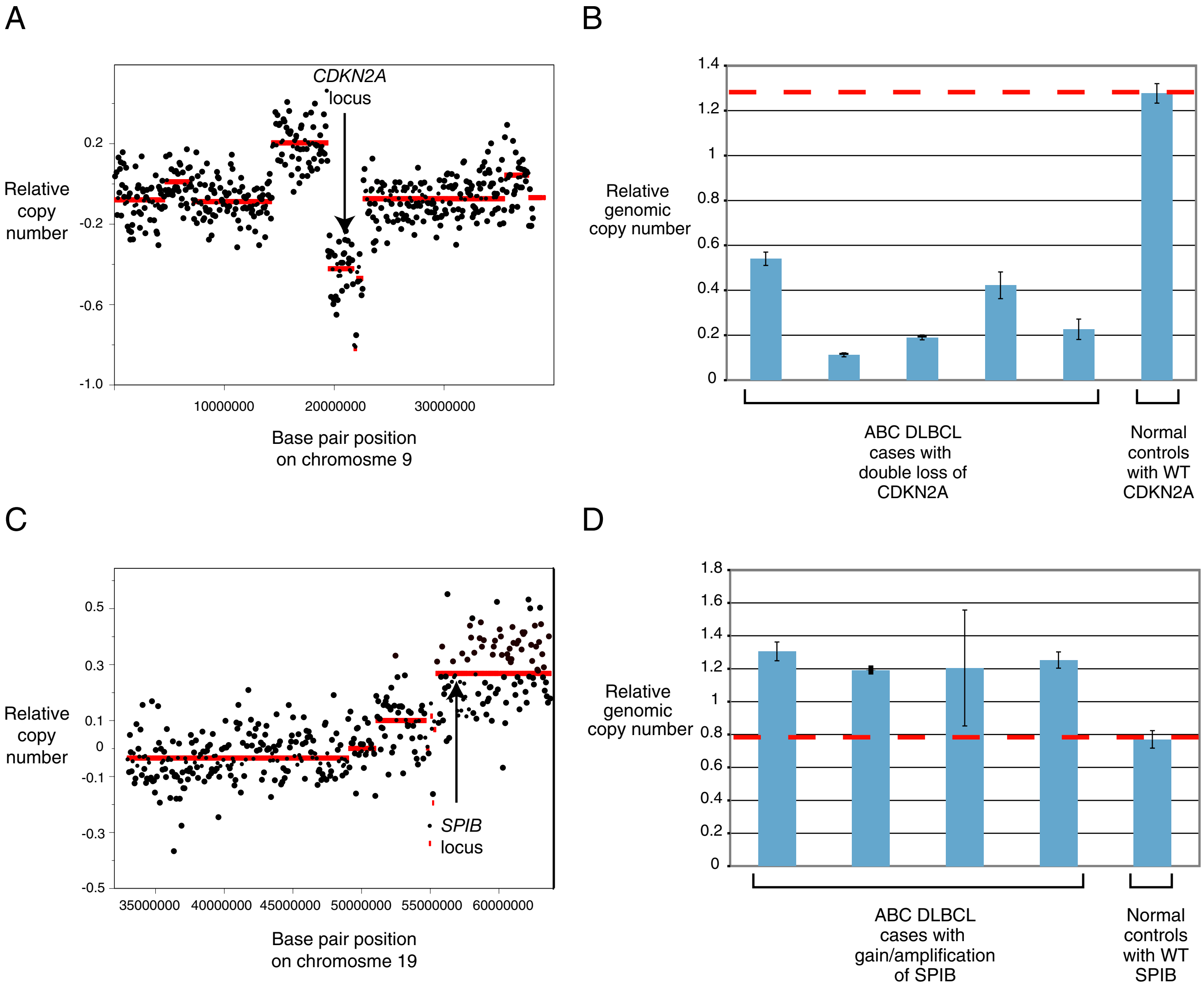

Fig. S3. (*A*) Deletion of *INK4a/ARF* locus detected by aCGH. (*B*) Quantitative real-time PCR of *INK4a/ARF* locus. Samples with homozygous deletion of *INK4a/ARF* locus have significantly lower relative genomic copy number than normal controls with wild-type locus. (*C*) Amplification of *SPIB* locus detected by aCGH. (*D*) Quantitative real-time PCR of *SPIB* locus. Samples with gain/amplification of *SPIB* locus have significantly higher relative genomic copy number than normal controls with wild-type locus.

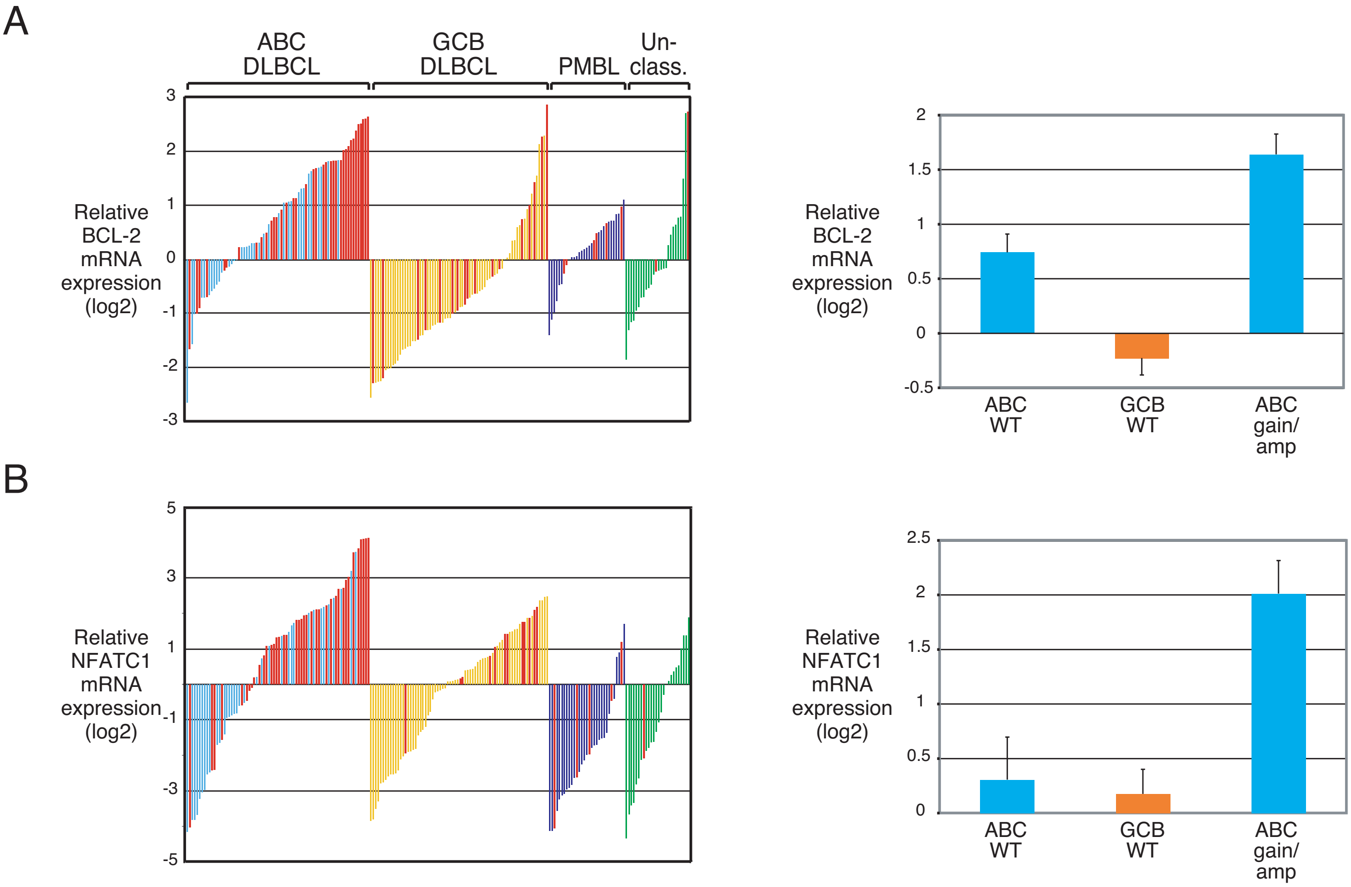

Fig. S4. Candidate oncogenes and tumor suppressors in DLBCL identified by aCGH. The left panels show the expression levels of the candidate gene in each case, with red bars indicating cases with the aberration. The right panels show the average expression of the candidate gene in cases grouped as indicated. (*A*) and (*B*) *BCL2* and *NFATC1* are candidate oncogenes associated with gain/amplifications of 18q in ABC DLBCL

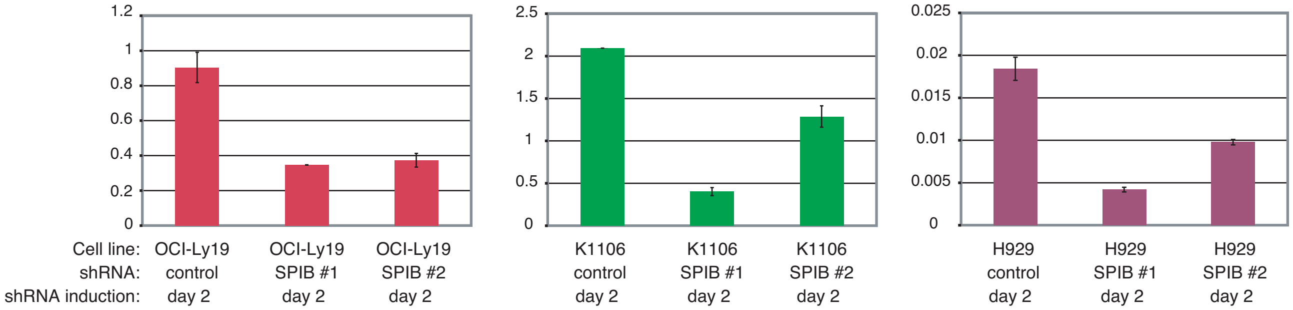

Fig. S5. Quantitative PCR analysis of *SPIB* mRNA 48 h following induction of SPIB shRNAs in indicated cell lines.

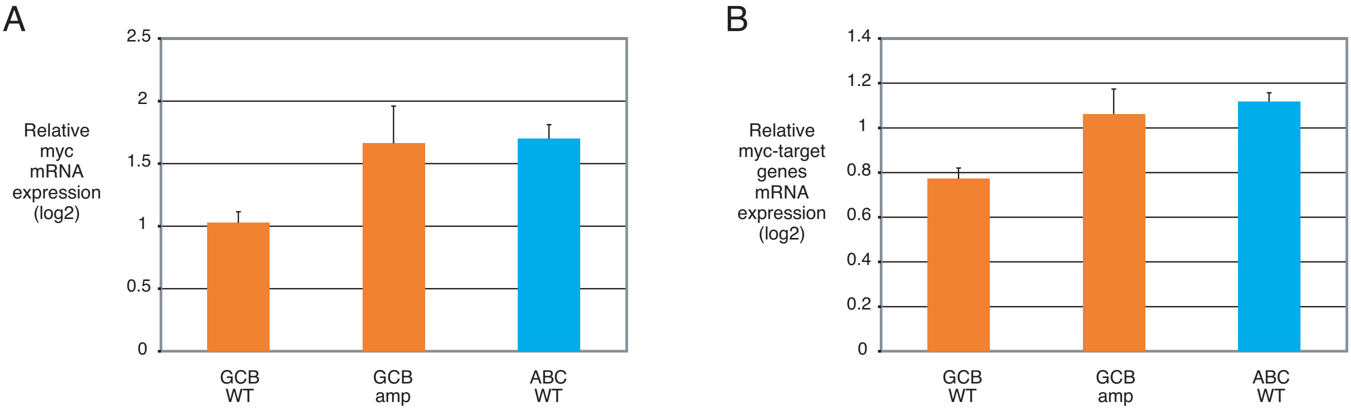

Fig. S6. (*A*) Average *myc* expression in cases grouped as indicated. (*B*) Average *myc* target gene signature expression in cases grouped as indicated.

Table S2. Subtype-specific MCRs in DLBCL detected by aCGH

	Chromosomal					DLBCL %‖			Subtype	Subtype
Type of aberration	location*	Start core†	End core‡	MCR peak§	*n*¶	ABC	GCB	PMBL	*P* value**	FDR††
Deletion	1p	0.82486	6.203951	2.436526	57	16.2	41.7	25.8	0.0028	0.033
Gain/amplification	1q	186.383798	197.90754	195.872682	21	6.8	20.8	0.0	0.0020	0.026
Amplification	2p	60.438887	62.010776	60.93399	32	5.4	27.8	19.4	0.0007	0.017
Gain/amplification	2p	60.438887	62.262464	60.966397	45	12.2	34.7	25.8	0.0049	0.048
Gain/amplification	Trisomy 3	0.0352	199.385052	n.a.	21	25.7	1.4	0.0	0.000001	0.0001
Gain/amplification	3p	0.0352	90.391956	n.a.	25	29.7	1.4	3.2	0.0000002	0.00003
Gain/amplification	3q	95.011055	199.385052	n.a.	35	39.2	2.8	6.5	0.00000001	0.000003
Gain/amplification	3q	100.802434	103.255067	103.060161	11	12.2	0.0	0.0	0.0008	0.016
Amplification	3q	102.081888	103.157085	103.060161	8	9.5	0.0	0.0	0.0054	0.049
Deletion	6q	62.025475	170.768683	n.a.	47	39.2	16.7	9.7	0.0008	0.016
Deletion	6q	67.067921	67.107365	67.085029	13	9.5	0.0	19.4	0.0006	0.015
Gain/amplification	Chromosome 9	0.002046	140.232212	n.a.	13	6.8	0.0	25.8	0.00004	0.0015
Gain/amplification	9p	0.002046	46.861416	n.a.	28	10.8	6.9	45.2	0.00002	0.0012
Amplification	9p	0.002046	46.861416	n.a.	12	2.7	1.4	25.8	0.00004	0.0015
Deletion	9p	21.683466	22.305714	21.925713	28	29.7	4.2	6.5	0.00003	0.0015
Double deletion	9p	21.801763	22.086442	21.925713	17	20.3	1.4	0.0	0.00005	0.0014
Deletion	9q	38.806775	70.144606	65.340023	40	31.1	19.4	3.2	0.0031	0.034
Gain/amplification	9q	65.340022	140.232212	n.a.	18	6.8	2.8	32.3	0.000063	0.0017
Deletion	Monosomy 10	0.084124	135.333576	n.a.	8	1.4	1.4	16.1	0.002	0.027
Deletion	10p	0.084124	39.114797	n.a.	10	2.7	1.4	19.4	0.0016	0.024
Deletion	10q	41.979103	135.333576	n.a.	8	1.4	1.4	16.1	0.002	0.029
Deletion	10q	84.498868	107.928835	89.618874	8	0.0	11.1	0.0	0.002	0.025
Gain/amplification	12q	63.054087	70.694251	64.663779	10	0.0	12.5	3.2	0.0014	0.026
Amplification	13q	90.176973	91.56868	90.807996	11	0.0	12.5	3.2	0.0014	0.025
Deletion	13q	109.212303	114.008196	111.028825	41	9.5	30.6	19.4	0.0053	0.050
Gain/amplification	16p	10.162907	11.685767	10.590398	11	12.2	0.0	6.5	0.0036	0.038
Gain/amplification	18q	16.778791	76.119357	n.a.	45	37.8	15.3	16.1	0.0041	0.042
Gain/amplification	19q	32.929831	63.795424	n.a.	17	18.9	4.2	0.0	0.0015	0.024
Gain/amplification	19q	54.658908	63.645106	58.213972	25	25.7	2.8	3.2	0.00003	0.0014
Gain/amplification	20p	0.00869	26.206155	n.a.	13	2.7	5.6	22.6	0.003	0.034

Type of aberration: type of MCR (deletion, double deletion, gain/amplification or amplification).

*Chromosomal location: chromosomal location of MCR.

†Start core: start of core MCR region in megabases.

‡End core: end of core MCR region in megabases.

§MCR peak: location of MCR peak in megabases.

¶n: absolute number of samples with MCR.

‖DLBCL%: percentage of samples in each DLBCL subtype (ABC DLBCL, GCB DLBCL, PMBL) with MCR.

**Subtype p-value: p-value of association between MCR and DLBCL subtype.

††Subtype FDR: false discovery rate of association between MCR and DLBCL subtype.

Table S3. MCRs associated with overall survival in DLBCL

Type of aberration*	Chromosomal location[†]	Start core[‡]	End core[§]	MCR peak[¶]	*n*[‖]	DLBCL (%)**			Hazard ratio	Survival *P* value[††]	Survival FDR[‡‡]
						ABC	GCB	PMBL			
Gain/amplification	3q	126.420189	126.657072	126.562637	16	12.1	4.2	6.5	3.17	0.0002	0.006
Amplification	3q	126.430781	126.657072	126.562637	8	6.8	2.8	3.2	4.91	0.0001	0.006
Gain/amplification	3p	0.0352	90.391956	n.a.	25	29.7	1.4	3.2	2.78	0.0001	0.006
Gain/amplification	3q	95.011055	199.385052	n.a.	35	39.2	2.8	6.5	2.7	0.00008	0.005
Gain/amplification	Trisomy 3	0.0352	199.385052	n.a.	21	25.7	1.4	0	3.55	0.00002	0.004
Deletion	9p	21.683466	22.305714	21.925713	28	29.7	4.2	6.5	2.68	0.00005	0.004
Gain/amplification	15q	69.894255	70.159074	70.02232	17	9.5	8.3	6.5	3.12	0.0001	0.006
Gain/amplification	18q	43.534181	76.119354	49.634237	9	10.8	1.4	0	5.09	0.00002	0.002

*Type of aberration: type of MCR (deletion, double deletion, gain/amplification, or amplification).

[†] Chromosomal location: chromosomal location of MCR.

[‡] Start core: start of core MCR region in megabases.

[§] End core: end of core MCR region in megabases.

[¶] MCR peak: location of MCR peak in megabases.

[‖] *n*, absolute number of samples with MCR

**DLBCL%: percentage of samples in each DLBCL subtype (ABC DLBCL, GCB DLBCL, PMBL) with MCR.

[††] Survival p-value: p-value of association between MCR and overall survival.

[‡‡] Survival FDR: false discovery rate of association between MCR and overall survival.

Table S4. Primers used for genomic real-time PCR

Gene	Primer name	Primer sequence
SPIB	SPIB_for	GCGTGAATGTCCCTTTGCA
	SPIB_rev	AAGCCCGGAGAAGACTCAGAT
INK4a/ARF	INK4a/ARF_for	CGCTGACCCTGGCAGTCT
	INK4a/ARF_rev	CAGCGATCTCTTGCACAAGTTT
B2M	B2M_for	CAGGATAAAGGCAGGTGGTTACC
	B2M_rev	TGGAAATGGCAGAAGAAAGATCA
PRKCQ	PRKCQex3A_for	CCTGGGACAGCACTTTTGATGC
	PRKCQex3A_rev	CACGGTGGTTTCAGAGATGAGGTC

Other Supporting Information Files

Table S1